



Effects of tonal alignment on lexical identification in Italian

Mariapaola d'Imperio, Caterina Petrone, Noël Nguyen

► To cite this version:

Mariapaola d'Imperio, Caterina Petrone, Noël Nguyen. Effects of tonal alignment on lexical identification in Italian. Carlos Gussenhoven. Tones and Tunes, Mouton de Gruyter, pp.79-106, 2007. hal-00150242

HAL Id: hal-00150242

<https://hal.science/hal-00150242>

Submitted on 29 May 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Effects of tonal alignment on lexical identification in Italian

Mariapaola D'Imperio, Caterina Petrone and Noël Nguyen

Abstract

In Neapolitan Italian, the LH nuclear accent rise is later in questions (L*+H) than in statements (L+H*). Furthermore, the H target is later in closed syllables than in open (penultimate) syllables, in both intonation modalities. As a consequence, the H target is differently located in lexical minimal pairs contrasting in consonantal length and syllable structure, such as *nonno* /'nɔnnɔ/ 'grandfather' and *nono* /'nɔno/ 'ninth': it is aligned towards the coda consonant in *nonno* and around the end of the stressed vowel in *nono*.

In this paper, we tested the hypothesis that Neapolitan listeners might employ such differences in alignment due to syllable structure in lexical identification. In Experiment I, listeners were asked to identify stimuli manipulated both in segmental length and tonal alignment as either *nonno* or *nono*. In Experiment II, listeners had to categorize stimuli manipulated in segmental length as well as for tonal alignment 1. as *nonno* or *nono* and, at the same time, 2. as questions or statements. Our results showed that, in stimuli ambiguous for segmental length, more *nonno* embedded in questions were perceived when the peak alignment was later within the coda consonant. This does not support models which predict that suprasegmental and segmental levels are independent from each other for the purpose of phonological contrast.

1. Introduction

A big debate related to the traditional distinction between the segmental and the suprasegmental levels of speech is whether such levels interact or are independent of each other. For instance, according to one of the most influential models of speech production, i.e., Levelt's (1989, 1993), each processing component is autonomous, in that it functions without taking into consideration the mode of operation and output of the other components. Specifically, within the Formulator component, the Grammatical Encoder accesses lexical items on the basis of semantic and syntactic constraints. Subsequently, the Phonological

Encoder provides the 'phonetic plan' for each lemma, i.e., how the item must be realized in the utterance. Prosody, as an aspect of speakers' articulation, is generated at this stage of processing. In perception, word recognition relies almost exclusively on the phonemic level. For example, according to the probabilistic connectionist model (Gaskell, Hare, and Marslen-Wilson, 1995), lexical identification can be modeled by a trained network exhibiting different degrees of lexical abstractness and phonological inference. Among these models, Norris, McQueen, and Cutler (2000) propose a speech recognition architecture merging lexical and phonemic information, in which the autonomy of the two levels is preserved by an additional phoneme decision stage. More recently, Stevens (2002) has claimed that lexical items are represented by sequences of segments, which are in turn composed of bundles of binary distinctive features. Words are then identified by extracting probability estimates for features values at the acoustic landmarks. Another recent proposal is that lexical entries are sparsely specified and that words are identified on the basis of feature mapping (Lahiri and Reetz 2002).

A common element in these models is that they do not predict any interaction between fundamental frequency (henceforth F_0) and the segmental level, since F_0 , being the main cue of intonation, is confined to the suprasegmental level. The existence of F_0 effects on segmental properties has been studied by relatively few investigations. Among them, Ladd and Silverman (1984) showed that differences in vowel intrinsic pitch might be accounted for by both segmental and prosodic effects. Silverman (1987, 1990) and Kohler (1990) found that listeners are capable of computing both macro- and microprosodic information in order to recover segmental structure, and that such microprosodic effects can influence segmental identification in a language-specific way. Moreover, Pierrehumbert and Talkin (1991) demonstrated that the English realization of the glottal stop and /h/ strongly depends on word and phrase-level prosodic effects. It must also be noted that an F_0 contour by itself is not always sufficient to convey an intonational meaning. For instance, Scherer, Ladd and Silverman (1984) found that the interpretation of intonation contours is compromised by low-pass filtering speech, which removes segmental information.

Another important finding is that the temporal location of tonal targets relative to segmental 'anchors' might be systematic (Arvaniti, Ladd, and Mennen 1998; Ladd, Faulkner, Faulkner, and Schepman, 1999). Specifically, in a variety of languages it has been found that the L(ow) tones of LH prenuclear rises are consistently aligned with the left edge of the stressed syllable (Caspers and van Heuven, 1993; Prieto, van Santen, and Hirschberg 1995; Arvaniti et al., 1998). However, the alignment of the H(igh) tones is much more controversial:

it can variably occur with the middle or the end of the stressed syllable, or even with the following unstressed syllable, as in the case of Greek (Arvaniti et al., 1998).

Tonal alignment is also sensitive to many phonetic and phonological factors (Silverman and Pierrehumbert, 1990; van Santen and Hirschberg, 1994; Rietveld and Gussenhoven, 1995; Ladd et al., 1999). Among these factors we find segmental length and syllable structure. For instance, Ladd, Mennen, and Schepman (2000) showed that the alignment of the peak of rising pitch accents in Dutch is affected by the phonological length of the accented syllable: Regardless of the actual acoustic duration of the vowels, accent peaks were aligned earlier in syllables containing phonologically long vowels than in syllables containing phonologically short vowels. This led the authors to argue that the variability of the H alignment in Dutch can be partly explained in terms of an anchoring of the H target to the right edge of the syllable (that is, at the right edge of the stressed vowel, when the syllable contains phonologically long vowels, or in the following consonant, when the syllable contains phonologically short vowels).

In this paper, we are concerned with tonal alignment in the Neapolitan variety of Italian. We know that in Neapolitan Italian, the nuclear rise LH as well as the following phrasal L- are systematically later in yes/no questions than in narrow focus statements (D'Imperio 2000, 2001, 2002; D'Imperio and House, 1997). Questions are characterized by a L*+H accent, with a peak around the end of the stressed syllable whereas statements are characterized by a L+H* accent, with a peak around the middle of the stressed vowel. The alignment difference of the LHL configuration appears to be systematically employed by Neapolitan listeners in order to recover the meaning contrast (D'Imperio, 2000; D'Imperio and House, 1997).

We also know that even though vowel duration is not contrastive in Italian, stressed vowels in open syllables are longer than stressed vowels in closed syllables, particularly in the penultimate position of the phrase (D'Imperio and Rosenthal, 1999; Van Santen and D'Imperio, 1999). Moreover, D'Imperio (2000) found that, similarly to the Dutch case, tonal alignment in Neapolitan Italian varies with syllable structure and vowel length. Both the L and the H tones of LH nuclear rises and the L- phrase accent (therefore, L1, H, and L2) are aligned later in closed syllables than in open syllables. As a consequence, the H target of the L*+H accent is differently located in lexical minimal pairs contrasting for consonantal length and syllable structure such as *pala* 'shovel' vs. *palla* 'ball' or also *note* 'note' vs. *notte* 'night'. In these minimal pairs, the first word contains a singleton consonant (which is represented orthographically by a single letter), whereas the second word contains a geminate consonant (which

is represented orthographically by a double letter). In fact, consonantal length is phonological in Italian. Furthermore, the vowel within the stressed (open) syllable of the word containing a singleton consonant is phonetically longer than that of the vowel within the stressed (closed) syllable of the word containing a geminate consonant. The result is that, in Neapolitan Italian questions, the peak occurs around the end of the stressed vowel in the word containing a singleton, whereas it tends to occur within the coda consonant in the word containing a geminate.

Our hypothesis is that Neapolitan listeners might capitalize on differences in alignment due to syllable structure for the identification of lexical minimal pairs contrasting in consonantal length and syllable structure. Specifically, our hypothesis is that Neapolitan listeners might identify more geminate items when the F_0 peak occurs within the coda consonant, since this alignment is congruent with closed syllable structure. In order to test this prediction, a first experiment (Experiment I) was carried out, in which Neapolitan listeners were asked to identify stimuli manipulated as to segmental length as well as tonal alignment. The listeners were asked to identify the items as either one of the lexical minimal pair *nonno* 'grandfather' vs. *nono* 'ninth'.

An additional experiment (Experiment II) was performed, in which two goals were pursued. First, since we know that gemination is mainly cued by consonant and stressed vowel duration (Esposito and Di Benedetto, 1999), we hypothesized that if tonal alignment affects the lexical identification, this effect would be stronger when segmental duration is ambiguous. Moreover, we asked whether variation in alignment between statements and questions might influence lexical identification. We know that effects of syllable structure can render the alignment of statement peaks similar to that of question peaks. Specifically, the H target of an L+H* (narrow focus statement) accent is usually realized near the stressed vowel offset when the vowel occurs in a closed syllable. This alignment is similar to that of an H target within a L*+H question pitch accent that is associated to an open syllable (D'Imperio, 2000). Though syllable structure effects do not clearly affect the identification of the statement/question contrast, one cannot exclude that this potential ambiguity might affect lexical identification. Hence, the second goal of this experiment was to verify if lexical identification might be affected by the interaction between the statement/question alignment contrast and the open/closed (syllable) alignment contrast.

2. Experiment I

2.1. Methods

2.1.1. Stimuli preparation

Two natural utterances were selected from a corpus of read sentences (D'Imperio, 2000), which were judged representative of the open/closed contrast within yes/no questions. These sentences were uttered by a native speaker of Neapolitan Italian (the first author) by producing an L*+H nuclear accent on the test words, as shown in Figure 1 (the stressed syllable is shown in capital letters):

- (1) *Vedrai il NONno?* 'Will you see the grandfather?'
- (2) *Vedrai il NOno?* 'Will you see the ninth?'

The two sentences constituted the basis for the perceptual manipulations for both Experiment I and Experiment II. The target word *nonno* /nɔnno/ (first sentence) contains a geminate consonant, and its stressed syllable is closed (*non-*); the second target word *nono* /nɔno/ (second sentence) contains a singleton consonant, and its stressed syllable is open (*no-*). For both test words, we measured (1) the duration of the stressed vowel and of the following consonant; (2) the alignment of the H tonal target relative to vowel offset.

Standard criteria of segmentation were used for the duration measurements (Peterson and Lehiste, 1960). Three segmental boundaries were marked: onset of the stressed vowel (v0), offset of the stressed vowel (c1) and offset of the following coda consonant (v1). Stressed vowel onset was located at the temporal point in which high amplitude periodicity started; boundaries between the stressed vowel and the following consonant were placed at the point in which a reduction in amplitude and a change in formant structure occurred. Coda consonant offset (v1) was placed after the nasal burst [ʔ] both for *nono* and *nonno*. In the case of *nonno*, such a criterion was chosen since there is no largely accepted principle for the segmentation of geminates (D'Imperio, 2000).

Duration measurements included the duration of the stressed vowel and of the following consonant for both syllables within the target words. In *nonno*, stressed vowel duration was equal to 146 ms while geminate duration was about 159 ms. In *nono*, stressed vowel duration was 178 ms, while the duration of the singleton [n] was 61 ms. Regarding tonal alignment measurements, we marked the temporal location of the F₀ maximum (H) within the nuclear L*+H pitch accent. In the case of *nono*, the peak was aligned close to the vowel off-

set, while for *nonno* the peak was aligned within the coda consonant (60 ms after the vowel offset).¹

In order to perform our perception experiments, both sentences were manipulated in two ways using the PRAAT speech analysis software (Boersma and Weenink, 2005; Wood, 2005): (1) manipulation of segmental duration; (2) manipulation of tonal alignment.²

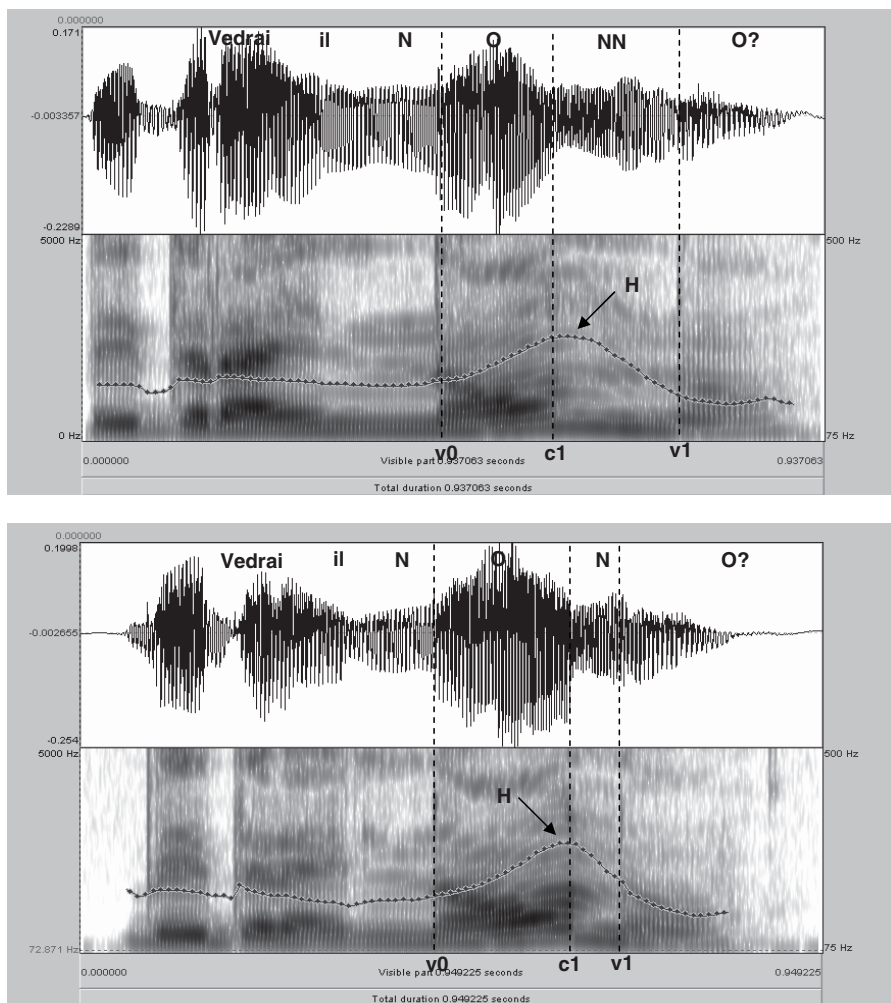


Figure 1. Waveform, spectrogram and f_0 curve for the two sentences *Vedrai il NONno?* 'Will you see the grandfather?' (upper panel) and *Vedrai il NONo?* 'Will you see the ninth?' (lower panel). Note the later peak alignment, with respect to the vocalic offset, in the test word containing a geminate.

Work on gemination has found that the manipulation of consonant duration is partly compensated for by the duration modification of the preceding stressed vowel (Esposito and Di Benedetto, 1999). Hence, in Experiment I, we first modified the proportion between the duration of the stressed vowel and the following consonant in seven steps so as to shift the percept from *nonno* to *nono* and *vice versa*. The specific manipulations were as follows: in *nonno* we progressively lengthened the duration of the stressed vowel in 7 steps of around 21 ms (duration of the whole vowel/7); at the same time, the following consonant was modified in an inverse proportional way. In *nono*, we progressively shortened the duration of the stressed vowel in 7 steps of around 26 ms (duration of the whole vowel/7); at the same time, the duration of the following consonant was lengthened in an inverse proportional way. In this way, we obtained a total of 14 stimuli, 7 from the closed syllable series (included the natural stimulus *nonno*) and 7 from the open syllable series (included the natural stimulus *nono*). Small differences in segmental duration between the two series should not be a matter of concern for this study, since it appears that the perception of gemination relies on the *relative* duration of the stressed vowel with respect to the following consonant and not on absolute duration values.

Subsequently, the temporal values of the L and H targets as well as the L target of the following L- phrase accent were modified. For all the stimuli obtained through the duration manipulation, a linear stylization ('t Hart, 1991) of the F_0 contour was carried out in which five points were interpolated: a point at the utterance onset; three points, respectively at the first L (L1), the H peak (H) and the second L (L2) location within the rising-falling contour; and a point at the utterance offset (see Figure 2). The F_0 value in Hz of the utterance onset as well as of the L1 target were both fixed at 210 Hz in order to avoid any "head level" effect. D'Imperio (2000) already employed this "head" value since it is intermediate between that of a typical question prenuclear region and a statement prenuclear region for the speaker (which is the same as in that study). The F_0 values for L1, H and L2 were taken from actual values within the base stimuli and were fixed at 210, 280 and 180 Hz, respectively. Finally, the utterance offset F_0 value was fixed at 150 Hz, in order to reproduce the F_0 phenomenon of declination. These fixed F_0 values allowed us to control for possible F_0 height effects.

Finally, the three tonal targets L1, H and L2 were simultaneously shifted leftwards or rightwards in time, relative to the vowel offset, in four steps of 15 ms (T1, T2, T3, T4). We also ensured that the alignment manipulations did not shift the percept of a yes/no question to that of a statement by taking tonal alignment values at which Neapolitan listeners generally perceive questions (D'Imperio and House, 1997; D'Imperio, 2000). In fact, our goal was simply to

create question patterns that would be more or less congruent with the syllabic structure of the base and not to intentionally shift perceived utterance modality.

In Table 1, F_0 target values and timing from c1 (stressed vowel offset) are reported. The tonal alignment manipulation is illustrated below (Figure 2).

Table 1. F_0 target values and timing relative to c1 (vowel offset) for L1, H, L2. T1, T2, T3 and T4 represent the four steps of tonal alignment manipulation

Tone Target	F_0 value (Hz)	Tonal alignment from c1 (ms)			
		T1	T2	T3	T4
L1	210	-135	-120	-105	-90
H	280	-15	0	+15	+30
L2	180	+105	+120	+135	+150

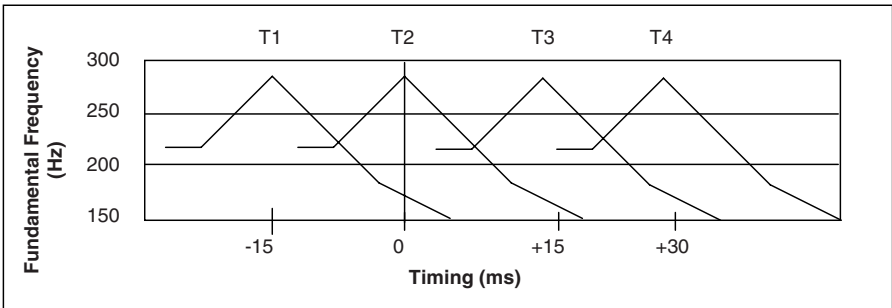


Figure 2. Schematic representation of the four tonal alignment steps T1, T2, T3, T4 relative to c1 (vowel offset).

All stimuli created by the combination of all the different steps of duration and alignment were resynthesized by means of PSOLA (*Pitch Synchronous Overlap and Add*, Moulines and Charpentier, 1990). Even though PSOLA allowed us to obtain natural sounding stimuli, we excluded 16 stimuli from this set, since they sounded quite unnatural and therefore might have biased our listeners' responses. Among these 16 stimuli, 8 were obtained from the combination of the two last steps of duration from the *nonno* base and the four steps of tonal target alignment; the other 8 discarded stimuli were obtained from the combination of the two last steps of duration from the *nono* base and the four steps of alignment. Intensity differences among the remaining 40 stimuli were normalized. The resulting set of 40 stimuli constituted the basis for our perception experiments.

2.1.2 Task and analysis procedure

The stimulus set was played directly from a laptop computer by means of PERCEVAL, software for performing computerized auditory and visual perception experiments (André et al., 1995-2003) developed at the Laboratoire Parole et Langage (Aix-en-Provence, France). The task was preceded by a training phase consisting of 10 trials selected from the stimulus set as being representative of both lexical categories *nonno* and *nono*. All the stimuli were presented binaurally through headphones in a studio at the Federico II University of Naples. At the same time, the two test words were visualized on the computer screen in the same order (1. *nonno*, 2. *nono*) to avoid uncertainty in responses due to order shifts.

The stimulus group was played six times in six different blocks with each stimulus occurring once per block. All the stimuli, as well as the practice trials, were randomly presented to avoid order effects. The start of each block presentation was preceded by a visual message on the screen. Moreover, the stimuli were all separated by a four-second pause; a sixty-second pause followed the end of each block. Listeners could also take a few-minute break between the third and the fourth block. The listeners were seated in front of the computer and were instructed to perform a two-alternative forced choice task, in which they had to identify the stimulus heard as either *nonno* or *nono*. They indicated their choice by clicking the right arrow on the keyboard for *nonno* and the left one for *nono*. All the listeners were right-handed. The experiment lasted about 25 minutes.

2.1.3. Participants

Thirteen listeners participated in the experiment, four females and nine males. The listeners, who were not paid for their participation, were all brought up in Naples and spoke standard Italian with a Neapolitan accent. All the participants were between 20 and 30 years old. None of them was a student in linguistics.

2.2. Results

The results of the identification task for stimuli modified from the closed-syllable base and from the open-syllable base are illustrated respectively in the upper and in the lower panel of Figure 3.

In both panels, percentages of *nonno* responses for all subjects (y axis) are shown across the five steps of duration manipulation (x axis). The four lines within each graph show responses for each of the four tonal alignment steps (T1, T2, T3 and T4). In the upper panel we see a decrease from *nonno* to *nono*

judgments as a function of duration manipulation in an orderly way: at early locations within the duration continuum (T1) *nonno* responses were dominant (100% *nonno* responses), while at the opposite end of the continuum *nono* responses were dominant (0% *nonno* responses), as expected from the literature on duration effects on geminate identification. The calculation of the crossover point showed that the category boundary between the two lexical categories was between D2 and D3 (mean boundary = 2.40). Moreover, no effects due to the tonal alignment manipulation were found between *nono* and *nonno* at duration steps D1, D3, D4 and D5. However, at D2 (vowel duration: 168 ms; consonant duration: 139 ms), we found a significant shift. Specifically, we found a higher percentage of *nonno* responses at early timing locations (T1:

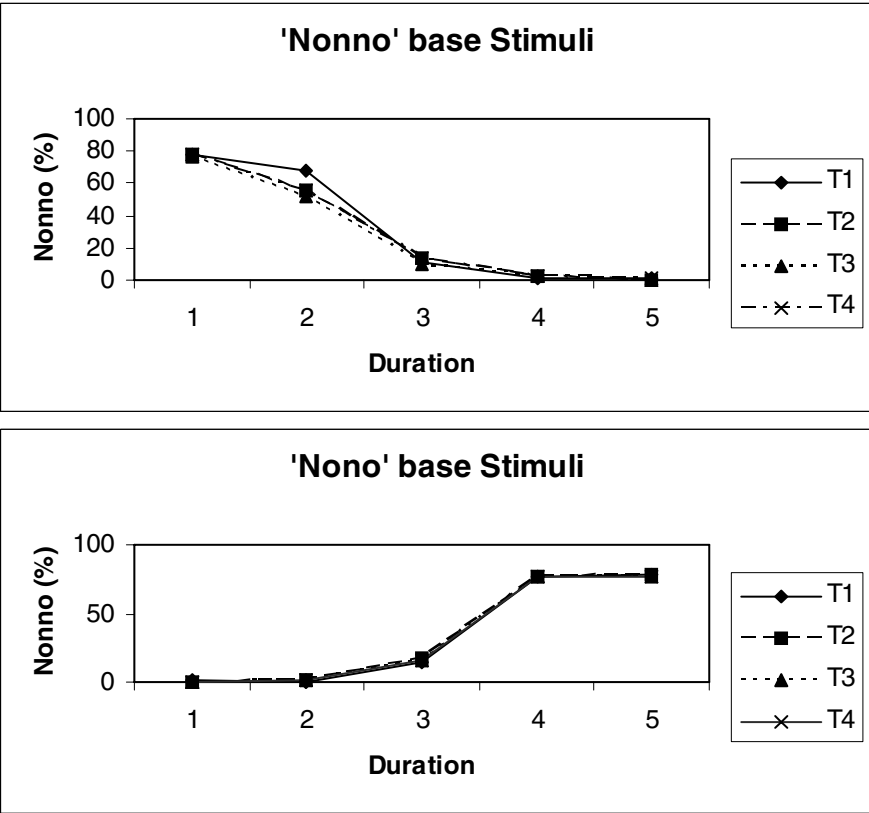


Figure 3. Percentage of *nonno* responses for the *nonno* base stimulus series (upper panel) and for the *nono* base stimulus series (lower panel) for all subjects. Note the shift due to tonal alignment in the upper panel at D2 duration step.

87.2%); this percentage becomes lower as the alignment becomes later (T2: 71.8%; T3: 66.7%), with a final slight increase at T4 (71.8%).

Analogously to the *nonno* series, in the *nonno* series (Figure 3, lower panel) lexical identification was strongly affected by the duration manipulation. In fact, *nonno* responses were dominant at the beginning of the duration continuum (0% *nonno* responses) while *nonno* responses were dominant at the end of the continuum (100% *nonno* responses). However, in contrast with the results of the *nonno* base series, no tonal alignment effects were found here.

A series of statistical analyses was performed to evaluate the effect of tonal alignment in lexical identification. Since an effect of tonal alignment was found only for stimuli at duration step D2 and only in the *nonno* base series, we only focused on these stimuli.

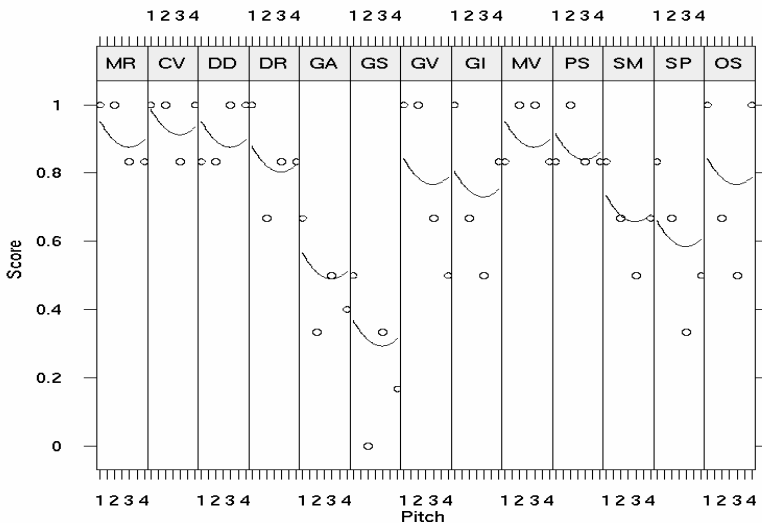


Figure 4. Polynomial fit by subject for *nonno* identification at D2 duration step, from the *nonno* base series. Mean *nonno* responses are plotted against timing steps. Note that 1, 2, 3, 4 in the x axis correspond respectively to T1, T2, T3, T4 steps.

First, we verified if response patterns could be linearly modeled. The linear model showed a good fit to the data [$p < .01$]. Moreover, a repeated measures ANOVA with Timing Step as between-items factor showed a globally significant effect of tonal alignment [$F(3,36) = 4.492$; $p < .01$]. We also asked whether the data might be described by a quadratic polynomial function. Results for the polynomial regression revealed that the model fitted the data well [$p < .01$].

Again, the effect of the tonal alignment manipulation was globally significant [$F(2,37) = 6.8870$; $p < .01$].

However, the ANOVA showed that the quadratic fitting did not obtain better results than the mere linear fitting, though the p -value was very close to the cutoff point [$p = .0163$]. In Figure 4, mean *nonno* responses are plotted for the four timing steps T1, T2, T3 and T4. Note that a tendency to identify more *nonno* items when the peak was aligned 30 ms after the vowel offset has been found for 8 subjects out of 13.

2.3. Discussion

The results of Experiment I appear to confirm the findings of the production and perception literature on geminate perception in Italian, i.e., consonant duration and preceding vowel duration are the main cues to distinguish lexical minimal pairs contrasting in consonantal length and syllable structure (Esposito and Di Benedetto, 1999). Moreover, in our experiment we also found a significant effect of tonal alignment, supporting the hypothesis that this cue might also be employed by Neapolitan listeners to identify lexical items. Namely, in stimuli with the same segmental duration (such as D2, in the *nonno* series), slight differences in peak alignment (which are usually found as an effect of longer vs. shorter vowel and absence vs. presence of coda consonant) shifted the category boundary between *nono* and *nonno*.

However, there are a number of issues to be raised regarding our data. First, it is noteworthy that a tonal alignment effect was found only for a subset of the stimuli relative to the second step of the duration manipulation (D2) in the *nonno* base series. One might hypothesize that tonal alignment information was employed to categorize such stimuli, because only in this case was segmental duration perceptually ambiguous. In other words, we believe that the duration step sizes employed in this experiment (except for step D2) were such that listeners continued to mainly employ segmental duration to categorize the stimuli. On the other hand, in the case of stimuli at D2, segmental duration was quite ambiguous, and therefore listeners were obliged to exploit other cues, such as tonal alignment, for lexical identification (see Figure 5).

Secondly, despite the significant effect of the tonal alignment manipulation, it is quite unclear just how listeners exploited this alignment information. We hypothesized that the later the alignment, the higher the percentage of *nonno* responses, since a late peak location should have been a better cue to closed-syllable structure type. However, in stimuli at D2 duration step from the *nonno* series, *nonno*-responses progressively decreased as tonal alignment values increased. This was true for T1, T2 and T3. Only when the peak was even later

within the geminate (T4), did we find a small increase of *nonno* responses for 8 listeners out of 13.

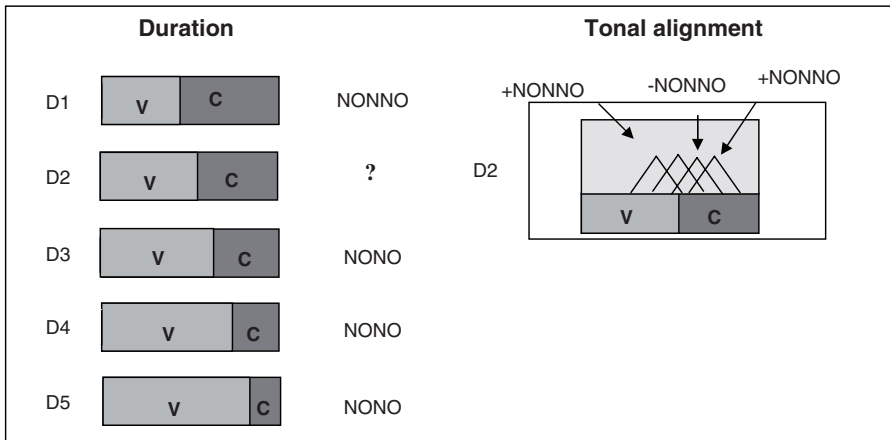


Figure 5. Schematic representation of the hypothesis that tonal alignment is mainly employed in stimuli that are ambiguous for segmental length (i.e., in D2 duration step from the *nonno* base series).

It could be argued that results found at early peak (T1 and T2) alignment might have been biased towards *nonno* by a potential ambiguity between the question and the statement meaning. We already know that in natural productions, the peak alignment in statements and questions can be very similar as an effect of syllable structure, even though statement/question identification does not seem to be affected by such acoustic ambiguity (see §1). However, one cannot dismiss the hypothesis that in our experiment the combined manipulations of tonal alignment and segmental duration might have shifted the category boundary between statement and question meaning, which might in turn have affected lexical identification. In other words, it is likely that when the peak is realized within the vowel (T1), listeners might have identified those stimuli as a *nonno* item embedded in a narrow focus statement instead of a *nono* item embedded in a yes/no question utterance. This ambiguity might have been even greater when the peak was aligned at vowel offset (T2).

What is more, the results relative to stimuli at T3 and T4 are also difficult to interpret. In fact, our results revealed that listeners identified more *nono* items when the peak was aligned 15 ms after the vowel offset (T3), while they identified more *nonno* items when the peak was 30 ms after this boundary. This would mean that perceptual targets are not strictly timed relative to acoustic landmarks, i.e., the vowel offset for *nono* and the following consonant for *nonno*, but are also computed relative to the actual duration of the vowels. We

know that in sentences uttered at normal speech rate, stressed vowels are longer in open syllables than in closed syllables, especially in penultimate position (Figure 1). Moreover, the F_0 peak is reached within the (phonetically) long vowel in open syllables, but it can be realized into the following coda consonant when the syllable is closed. However, when the stressed vowel (and coda consonant) duration is ambiguous between open and closed syllable items, listeners would appear to employ a constraint referring to the duration of the rise, so that the early location of the H target within the coda consonant might be a cue for an open syllable.

On the other hand, in the same ambiguous stimuli, the later alignment of the peak within the coda consonant might be employed as a cue for *nonno*. That is, differences in early (as in T3) vs. late (as in T4) alignment within the coda consonant are taken as evidence for the structural contrast between open vs. closed syllables. We hypothesize, then, that the perceptual target for *nono* is computed early in the coda consonant in stimuli that are ambiguous as to their vowel and consonant duration. We also expect that when the peak is aligned later within the same coda consonant, listeners can perceive more *nonno* than *nono*.

Experiment II was conducted to test the hypotheses that lexical identification is affected by tonal alignment when the durational composition is ambiguous and that it interacts with the perception of intonational meaning.

3. Experiment II

The first hypothesis tested in Experiment II is that tonal alignment is systematically employed to distinguish lexical items contrasting in consonantal length and syllable structure when the main cue for such a contrast (i.e., segmental duration) is ambiguous. If this hypothesis is true, we expect a stronger effect of tonal alignment in highly ambiguous stimuli than in the stimuli employed in Experiment I. On the other hand, if it is not true, we expect that scores for the *nono/nonno* responses will be around chance for all ambiguous stimuli, independent of tonal alignment. Moreover, if this hypothesis is confirmed, we also expect that shifting the pitch peak later within the ambiguous coda consonant will progressively change the percept from *nono* to *nonno*, since later alignment within the coda consonant is more congruent with the usual alignment in closed syllables.

The second goal of Experiment II was to verify the existence of an interaction between the intonation meaning and the lexical contrast. Assuming that all the ambiguous stimuli with a peak aligned before the stressed vowel offset will

be categorized as statements, while all the ambiguous stimuli with the peak aligned after the vowel offset will be categorized as questions, we hypothesized that the late alignment of the pitch peak will score more *nonno* responses both in statements and in questions, since we know the H target is later in closed syllables than in open syllable in both modalities (Figure 6).

In order to obtain more precise information on the ambiguous durational composition of the *nono-nonno* continuum, we performed a preliminary task, described in the next section. This information will be used as a basis for our tonal alignment manipulations in Experiment II.

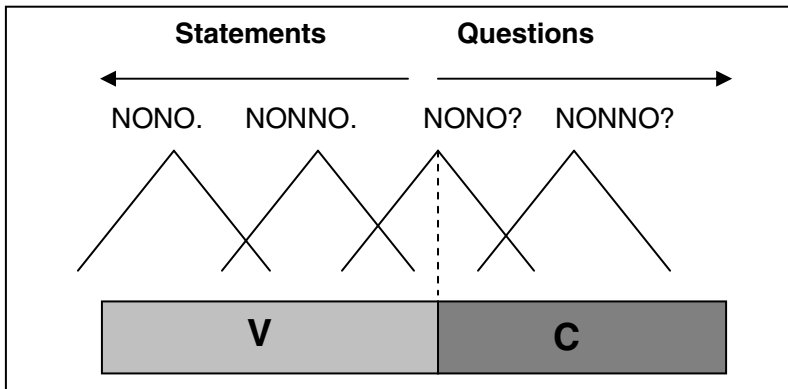


Figure 6. Schematic representation of the hypothesis that tonal alignment manipulation in Experiment II will affect both the intonation and the lexical identification: More statements will be identified when the H target is aligned before the vowel offset, while more questions will be identified when the H target is aligned after the vowel offset. At the same time, more *nonno* will be identified when the H target is later in the test syllable, in both intonation modalities.

3.1. Method

3.1.1. Preliminary task

Our aim was to find stimuli that are ambiguous as to the length opposition between *nonno* and *nono*, to be used for the tonal alignment manipulations. The results of Experiment I showed a crossover boundary between *nonno* and *nono* between the second (vowel duration: 168 ms; consonant duration: 139 ms) and the third (vowel duration: 189 ms; consonant duration: 120 ms) step of duration in the *nonno* stimuli series (Figure 3, upper panel). In the *nono* stimuli series (Figure 3, lower panel), the same boundary was between the third (vowel duration: 125 ms; consonant duration: 87 ms) and the fourth (vowel duration:

99 ms; consonant duration: 109 ms) duration step. Therefore, we decided to manipulate vowel and consonant duration only within these regions, since they appear to be crucial in the transition between the two lexical categories.

However, since we slightly changed the boundary location between the segments, we could not directly compare the stimuli created with those of Experiment I. Rather, we modified the two original test words, always embedded in yes/no questions, as shown in Figure 1. On the one hand, we modified the *nonno* original base in order to create a stimulus similar in duration to the one obtained for duration step D2 of Experiment I. On the other hand, we manipulated the *nono* base until we obtained a stimulus similar in duration to the one corresponding to duration step D3 of Experiment I.

These two stimuli were the basis of a progressive duration manipulation of the stressed vowel in 5 steps of 5 ms. The duration of the following consonant was also modified, in an inversely proportional way. Note that such a small step size for vowel manipulation allowed us to avoid the creation of stimuli whose percept suddenly changed from *nonno* to *nono* and vice versa.

In this fashion, we obtained 5 stimuli from the base stimulus used for the *nonno* series and 5 from the base stimulus used for *nono* series. Three stimuli from this set were eliminated, since we judged them perceptually unambiguous. Specifically, we excluded one stimulus, created from the last step of duration manipulation in the *nonno* series and two stimuli from *nono*, obtained from the last two steps of the duration manipulation in the *nono* series. The two stimuli employed as base for the duration manipulations as well as the natural stimuli were added to the corpus. Hence, we obtained a set of 11 stimuli, which constituted the basis of our preliminary task. Vowels and consonant duration values for both the *nonno* and the *nono* series are reported in Table 2.

This set of stimuli was presented to 5 Neapolitan listeners following the same procedure of Experiment I (see § 2.1). Participants were asked to listen to these stimuli and to categorize them as *nonno* or *nono*. This set, preceded by 6 practice trials, was randomly played 7 times from a laptop placed in a studio at the University of Naples. The results of the identification task, expressed as a percentage of *nonno* responses, are shown in Table 3.

As expected, vowel and consonant duration manipulation affected lexical identification in both series. Specifically, in the *nonno* base series the highest percentage of *nonno* responses was obtained at duration step D1 (i.e., the original stimulus) as well as at duration step D2 (i.e., the one used as a basis for our duration manipulations). The percentage dropped to 40%–50% of *nonno* responses at steps D3, D4, D5, D6. Analogously, in the *nono* series the percentage of *nonno* responses increased from step 1 up to the last one. However the

percentage was much lower at steps D2–D4, reaching 57% only in the last duration step (D5).

Table 2. Duration values for stressed vowel and following coda consonant in the *nonno*-base stimuli and in the *nono* base stimuli. In each group, stimuli at D1 duration step are the natural ones. Stimuli at D2 duration step are those used as basis of duration manipulation

<i>nonno base</i>	V duration (ms)	C duration (ms)
D1	146	160
D2	169	133
D3	173	128
D4	178	125
D5	183	121
D6	188	118
<i>nono base</i>	V duration (ms)	C duration (ms)
D1	178	61
D2	135	87
D3	130	90
D4	125	94
D5	120	98

Table 3. Percentage of *nonno* responses for both the *nonno* base stimulus series and the *nono* base stimulus series.

<i>nonno base</i>	<i>nono %</i>
D1	100
D2	71.4
D3	42.9
D4	40
D5	45.7
D6	48.6
<i>nono base</i>	<i>nonno %</i>
D1	0
D2	5.7
D3	5.7
D4	20
D5	57.1

One could ask whether differences in responses between the *nonno* base stimuli and the *nono* base stimuli could be accounted for by some property of the base stimulus, which could have affected the lexical identification favoring one answer over the other. At the moment, we cannot answer such a question. In this task we focused on the role of duration as the main correlate of gemination, without taking into consideration other potential (segmental and suprasegmental) cues. For the same reason, we did not explicitly control for tonal alignment information either, which may also have influenced lexical identification (but see Petrone & D'Imperio, in preparation [?]).

In sum, the results of the preliminary task show that the most ambiguous stimuli are produced at duration steps 5 and 6 (i.e., D5 and D6) of the *nonno* base series. Specifically, D5 scored 45.7 % of *nonno* responses, whereas D6 scored 48.7 % of *nonno* responses. Hence, both of these stimuli were employed as a basis for the tonal alignment manipulations of Experiment II.

3.1.2. *Experiment II: Stimuli preparation*

Following procedures as used for Experiment I, a linear stylization of the pitch contour was made for both stimuli obtained at D5 and D6 duration steps (i.e., the ambiguous stimuli). Specifically, we stylized: (1) the pitch contour of the sentence *Vedrai il nonno?*, containing the D5 stimulus; and (2) the pitch contour of the sentence *Vedrai il nonno?*, containing the D6 stimulus. We kept F_0 height at the same values of the stimuli in Experiment I. Once the stylization was applied, tonal alignment was modified.

In order to verify whether the statement/question contrast affects lexical identification, we simultaneously shifted the LHL configuration leftwards and rightwards to obtain an equal number of narrow focus statements and yes/no questions. All the shifts were made relative to the vowel offset, since it appeared from the results of Experiment I that vowel offset could be a perceptual boundary between statements and questions. We also ensured that tonal alignment manipulations did not produce unnatural sounding stimuli.

By this procedure, 14 alignment steps were created for both D5 and D6. Specifically, the three target locations L1, H and L2 were shifted in 7 steps up to the vowel offset in order to obtain statement utterances and they were shifted in 7 steps from the vowel offset onwards in order to obtain questions (Figure 7). The step size was 15 ms, as in Experiment I. Finally, the 28 stimuli, created by the combination of the two steps of duration and the sixteen alignment steps were resynthesized through PSOLA. This corpus constituted the basis for our perception experiment. The actual timing values are presented in Table 4.

Table 4. Tonal alignment values for L1, H, L2 relative to vowel offset.

Tone Target	Tonal alignment from C1 (ms)													
	Narrow focus statement							Yes/no question						
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14
L1	-225	-210	-195	-180	-165	-150	-135	-120	-105	-90	-75	-60	-45	-30
H	-105	-90	-75	-60	-45	-30	-15	0	15	30	45	60	75	90
L2	15	30	45	60	75	90	105	120	135	150	165	180	195	210

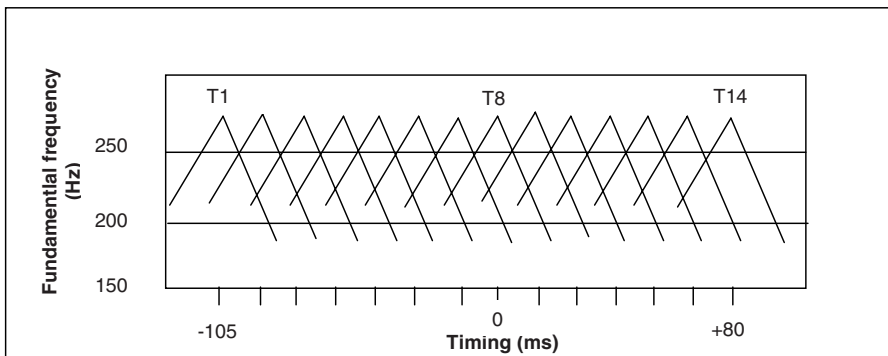


Figure 7. Schematic representation of the fourteen tonal alignment shifts from T1 to T14 relative to c1 (vowel offset). Note that, according to our hypothesis, all the stimuli with the peak aligned before the vowel offset (T1-T7) should be categorized as (narrow focus) statements whereas stimuli with the peak aligned at or after the vowel offset (T8-T14) should be categorized as (yes/no) questions.

3.1.3. Experiment II: Experimental procedure

The set of stimuli, preceded by 6 practice trials, was played through a laptop by means of the PERCEVAL software package. The stimulus set was played 7 times in 7 randomized blocks, with each stimulus occurring only once per block. A seven-second pause occurred at the end of each stimulus; the blocks were all separated by a sixty-second pause. All the stimuli, as well as the practice trials, were randomly presented to avoid order effects. The stimuli were presented binaurally via Sennheiser HD497 headphones in a studio at the University of Naples. Twelve Neapolitan listeners participated in the experiment. All the subjects were between 20–30 years of age and they were not paid for their participation. After listening to each stimulus, the subjects performed a

four-choice forced task, in which they had to categorize the stimulus as either *nonno* or *nono* and, at the same time, as a question or a statement. Listeners had to report their responses by pressing one of 4 buttons on a box (Ghio, André, Teston, and Cavé, 2003), which were labeled as follows: *NONNO?*, *NONNO.*, *NONO?*, *NONO.* These labels were also visualized on the computer screen, always in the same order. The box was directly linked to the laptop, so that subjects' responses were automatically stored in the computer memory.

3.2. Results

In Figure 8 the results are separately plotted for the intonation contrast (upper panel) and for the lexical contrast (lower panel). The upper panel shows the percentage of question responses pooled for all the listeners (y axis) across all time steps (x axis) for both base stimuli D5 and D6. The results for the intonation contrast confirmed data already known in the literature, in that the questions/statements identification is a function of the rise-fall timing. Namely, the perception of questions (which was the base of our manipulations) decreased by shifting the rise-fall leftwards in an orderly way, i.e., from the end to the beginning of the continuum. The results were similar for D5 and D6.

Here, we are particularly concerned with the results obtained for stimuli at T7, T8, T9 and T10. In fact, timing values for these steps corresponded respectively to those in T1, T2, T3, and T4 in Experiment I. Note that in T7 question responses are about 70% for both the stimuli at D5 and D6, whereas in T8 question responses reach 80% only for the stimulus at D6. This percentage is also reached in the following steps T9 and T10 for both base stimuli.

Let us now look at the results for the lexical contrast. In the lower panel of Figure 8 we report the percentage of *nonno* responses pooled for all the listeners (y axis) across time steps (x axis), for D5 and D6. The graph shows that *nonno* responses do not drastically vary as a function of timing. This was expected, since the main cue for the *nonno/nono* contrast (i.e., duration) is perceptually ambiguous in both D5 and D6. However, note that the percentage of *nonno* responses was affected by tonal alignment, as the responses decreased in stimuli categorized as statements (T2–T5) and increased in stimuli categorized as questions (T9–T13) in an orderly way.

Strikingly, the results at T2–T5 went in the opposite direction to the one expected. Remember that we hypothesized that late alignment would have scored more *nonno* responses. Here, on the contrary, in stimuli judged as statements, *nonno* responses were dominant at the beginning of the continuum, for both base stimuli (at T2, D5 and D6 scored respectively 71.4% and 66.7% of *nonno* responses). The percent of *nonno* suddenly decreased as the H target was

shifted towards the end of the vowel (at T5, D5 and D6 scored respectively 40.5% and 44% of *nonno* responses). Results for stimuli at T6, T7 and T8 are even more unexpected, since the percentage of *nonno* responses suddenly increased at T6 and, after a decrease at T7, increased again at T8, for both stimuli D5 and D6.

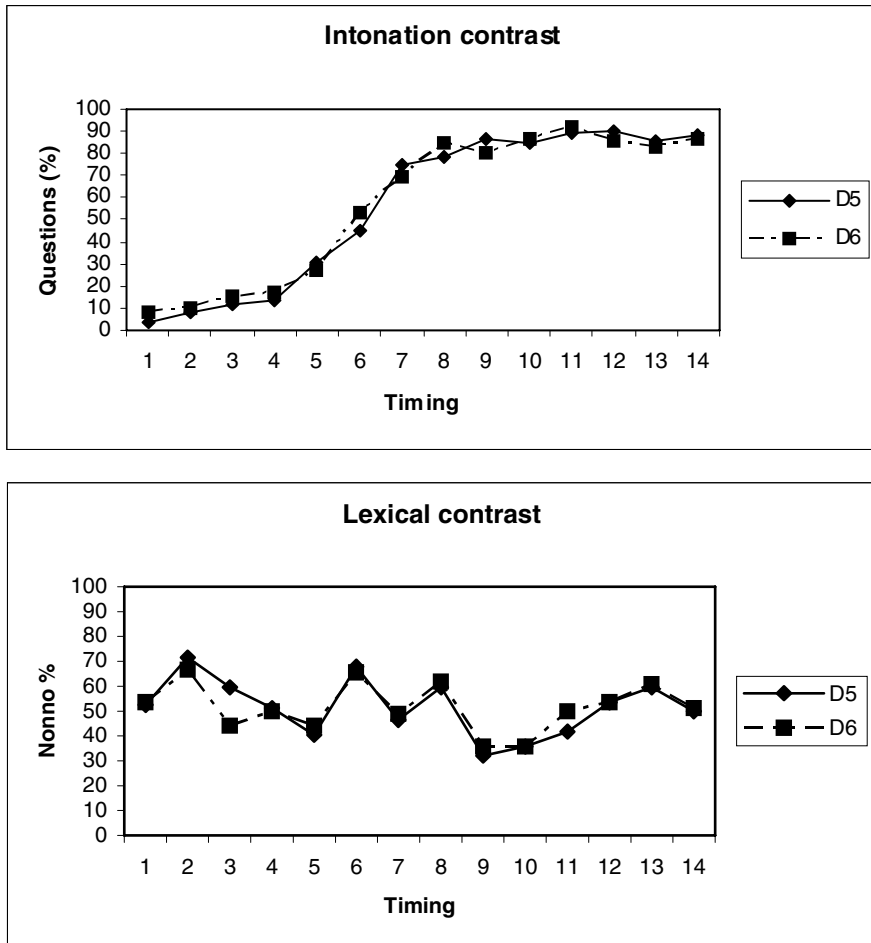


Figure 8. Percentage of questions responses for the question/statement identification task (upper panel) and of *nonno* responses for the lexical identification task (lower panel) for both ambiguous stimuli D5 and D6.

Finally, in stimuli clearly judged as questions (T9 – T14), *nonno* responses increased in an orderly way. This was expected, since our hypothesis was that in questions the later the alignment in the coda consonant, the higher the ob-

tained percentage of *nonno* responses. Specifically, when the peak was aligned at 15 ms after the vocalic boundary (T9), the lowest percentage of *nonno* responses was obtained both for D5 and D6 (32.1% and 37.7%, respectively). On the other hand, when the peak was aligned at 30 ms after the vowel offset (T10), we found a small increase of this percentage in D5 (in fact, both stimuli scored 35.7%). The percentage of *nonno* continued to increase, until it reached the highest value at T13 both for D5 and D6 (59.5% and 60.8%, respectively).

A series of statistical analyses was performed. We concentrated on data for steps T9 to T14, since these stimuli were clearly categorized as questions. In order to statistically analyze these results, data obtained at T9–T14 timing steps were smoothed using a logistic regression model. Results for the Analysis of Deviance showed that tonal alignment significantly affected *nonno* responses ($p < 0.01$). This confirms that shifting the peak rightwards within the coda consonant leads listeners to categorize such stimuli as *nonno*. In Figure 9, *nonno* responses percentages are plotted against timing steps from T9 up to T14, for both D5 and D6 and separately per each speaker. Note that in 6 subjects out of 12 there was a trend to identify more *nonno* with later peak alignment in the coda consonant. Hence, despite the variability observed within the data, tonal alignment appears to be employed as an active cue for lexical identification in Neapolitan Italian.

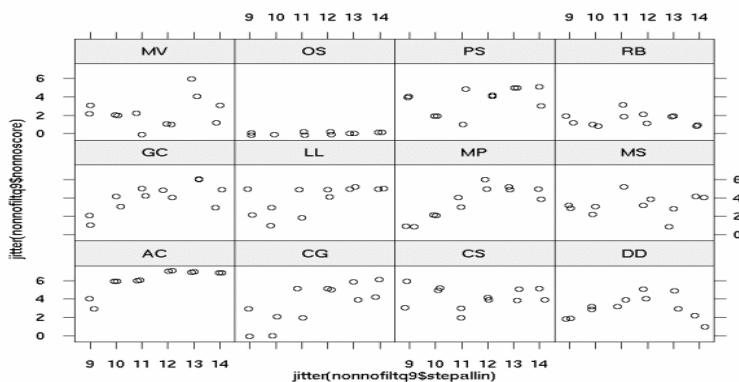


Figure 9. Dispersion plot for the *nonno* scores at T9–T14 timing steps. The results are reported for each subject. Note that for subjects GC, LL, MP, AC, CG, CS *nonno* scores increase as alignment is shifted rightwards.

3.3. Discussion

The results of Experiment II suggest that Neapolitans can systematically employ tonal alignment detail (due to segmental duration and syllable structure

effects) in order to identify lexical items that are ambiguous for segmental length. This does not support the idea that F_0 information is autonomous from the segmental one in lexical identification. In fact, if lexical identification had not been affected by tonal alignment, the score for the *nonno/nono* identification would have been around chance, independent of tonal alignment manipulations (see § 3).

However, the results for lexical identification for stimuli at T1–T8 are quite difficult to interpret. Specifically, for stimuli T1–T5, we found that more *nonno* are identified when the peak is aligned early within the stressed vowel, while our hypothesis was that more *nonno* would be identified as the alignment becomes later, both in statements and in questions. (see §1). At this point, it might be premature to draw a conclusion from these data. As mentioned above, D'Imperio (2000) found a significant effect of base utterance type in the statement/question identification, in that the crossover boundaries were later in the statement base series than in the question base series. This suggests the existence of correlates for the statement/question contrast other than tonal alignment. Similarly, one cannot exclude that in Experiment II lexical identification responses for stimuli judged as statements might be biased by some properties of the base stimulus employed for our manipulations (i.e., a question).

The results at T6–T8 are even more difficult to explain. We know that in questions the H target is aligned towards the vowel offset in open syllables, and towards the following coda consonant in closed syllable. Specifically, in closed syllables the peak can be aligned both around the vowel offset and at the beginning of the following consonant, depending on the speaker. Then, one explanation could be that when the peak was around the vowel offset (i.e., at T6–T8), lexical identification might have been affected by the ambiguity between an open and a closed syllable. Alternatively or additionally, lexical identification may have been affected by ambiguity in intonation meaning. A score of 80% correct responses usually is considered as a good criterion for successful completion of a test. If we require an 80% score for successful question/statement identification, then this criterion was not met at T6 and T7 timing steps either (for both D5 and D6), while at T8 questions responses were slightly more than 80% only for D6 base stimulus. This would mean that stimuli at T6–T8 were ambiguous between statements and questions. A more controlled study is needed to better evaluate the effects of syllable structure in narrow focus statements as well as the interaction between the statement/question contrast and the closed/open syllable contrast.

Here, we would like to concentrate on results for stimuli at T9–T14, since they were clearly judged as questions (i.e., our original base utterance). Data obtained from Experiment II appear to support the hypothesis that in stimuli

ambiguous for segmental length, differences in tonal alignment within the coda consonant can shift the percept of *nono* to *nonno*. Such results appeared to confirm the trend emerging from Experiment I. In Experiment I more *nono* were perceived when the peak was at 15 ms after the vowel offset (T1), but a small increase of *nonno* responses was found when the peak was 30 ms after the vowel offset. In Experiment II, these results are replicated by stimuli at steps T9 and T10, corresponding to steps T3 and T4 of Experiment I. Specifically, more *nono* were identified with stimuli at T9, both for the stimuli ambiguous for length (D5 and D6), whereas *nonno* responses slightly increased at T10 for D6 base stimulus. In Experiment II we also found that the percept of *nonno* increased as the peak alignment was later in the coda consonant, suggesting that listeners employ this detail as a cue for syllable structure.

The results in Experiment II appear to support the hypothesis, already formulated in Experiment I, that perceptual targets are not strictly anchored to acoustic landmarks, as we expect from the acoustic claims of the strict segmental anchoring hypothesis (Arvaniti et al., 1998; Ladd et al., 1999). In fact, if this were true, we would expect the H target to be stably aligned to the vowel offset for *nono* and to the beginning of the coda consonant for *nonno*, regardless of the segmental duration manipulation. On the contrary, participants to our experiments seemed to take into consideration durational constraints, (e.g., that the F_0 rise needs a certain amount of time to be completed) since the perceptual target for the H tone appears to be located after the vowel offset when vowel duration is shorter than expected for (phonetically) long vowels in open penultimate syllable.

On the other hand, a later peak in the coda consonant appears to better cue closed syllable structure, since the late alignment is more congruent with this syllable type. This means that the perceptual location of tonal targets strongly depends on syllable structure differences. In our view, our finding seems to perceptually validate the “revised structural explanation” claimed by Ladd et al. (2000) for Dutch, in that Neapolitan listeners are able to compute differences in alignment due to syllable structure, even though duration manipulation partly affected perceptual target location.

4. Conclusion

Despite the complexities presented by our data, the results of Experiment I and II have important consequences for the debate about the nature of tonal alignment. Atterer and Ladd (2003) found gradient differences in the alignment of prenuclear rises, both across varieties (Northern vs. Southern German) and

across languages (German vs. English). They therefore suggested the existence of a single category of rising accent, even though tonal target alignment can vary along a continuum across languages and varieties of the same languages due to language-specific phonetic implementation rules. We argue here, however, that in the case of Neapolitan Italian tonal alignment has phonological status, in that listeners systematically employ alignment cues in order to distinguish narrow focus statements (L+H*) and yes/no questions (L*+H). In this paper we have shown that differences in H target alignment due to phonotactic constraints might be employed for lexical identification of minimal pairs of Neapolitan Italian such as *nonno* and *nono*. In other words, the variability in H alignment due to syllable structure constraints appears to be exploited by Neapolitan listeners to convey non-intonational information. These results suggest that both strictly segmental and tonal alignment information might be perceptually relevant to the identification of lexical items contrasting in segmental length and syllable structure. That is, this phonological contrast might rely on cues other than mere segmental duration.

On the basis of our results, together with those obtained by Petrone and D'Imperio (forthcoming), we believe that suprasegmental and segmental information are not separated from each other for the purpose of phonological contrast. Our results thus do not support models such as Levelt's (1989, 1993), in which speech processing is the result of the independent contribution of different information sources and in which prosody is generated only when lexical access and syntactic building procedures are completed. Also, our results are in contrast with perception models in which word prosody is not taken into consideration for lexical identification.

Notes

1. The word-initial /n/ was 147 ms in *nonno* and 115 ms in *nono*, while the word-final vowel /o/ was 141 ms in both the target words. Note that no attempt was made in order to control also for word-initial consonant duration, since there is no clear data in literature showing the existence of within-word gemination effects on the duration of the consonant *preceding* the stressed vowel.
2. One assumption of this paper is that the only segmental cue for gemination is duration. Specifically, note that in a study on acoustic correlates of singleton vs. geminate consonants in Italian, Mattei and Di Benedetto (2000) found that duration is the main cue of nasal consonants gemination. Moreover, they found a small effect of gemination in consonant energy, in that the total energy of the nasal consonant increased in the geminate case. However, changes in nasal consonant energy due to gemination were not taken into consideration for our manipulations, since they do not seem perceptually relevant (Mattei, personal communication).

References

- André, Carine, Alain Ghio, Christian Cavé, and Bernard Teston
12003 *PERCEVAL: PERception EVALuation Auditive and Visuelle* (computer program), <http://www.lpl.univ-aix.fr/~lpl/dev/perceval>.
- Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen
1998 Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3–25.
- Atterer, Michaela, and D. Robert Ladd
2003 On the phonetics and phonology of “segmental anchoring” of F_0 : evidence from German. *Journal of Phonetics*, 32: 177–197.
- Boersma, Paul, and David Weenink
2005 Praat: doing phonetics by computer (computer program), version 4.3.01 <http://www.praat.org/>
- Caspers, Johanneke, and Vincent J. van Heuven
1993 Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161–171.
- D'Imperio, Mariapaola, and David House
1997 Perception of Questions and Statements in Neapolitan Italian. In *Proceedings of Eurospeech '97*, G. Kokkinakis, N. Fakotakis and E. Dermatas (eds.), vol. 1, 251–254. European Speech Communication Association: Rhodes, Greece.
- D'Imperio, Mariapaola, and Sam Rosenthal
1999 Phonetics and phonology of main stress in Italian. *Phonology* 66: 1–28.
- D'Imperio, Mariapaola
2000 The Role of Perception in Defining Tonal Targets and their Alignment. Ph. D. diss., Department of Linguistics, The Ohio State University.
2001 Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33 (4): 339–356.
2002 Italian Intonation: an overview and some questions. *Probus* 14 (1): 37–69.
- Esposito, Anna, and Maria Gabriella Di Benedetto
1999 Acoustic and perceptual analysis of Italian stop consonants. *The Journal of the Acoustical Society of America* 106 (4): 2051–2062.
- Gaskell, M.Gareth, Mary Hare, and William D.Marslen-Wilso
1995 A connectionist model of phonological representation in speech perception, *Cognitive Science* 19: 407–439.
- Ghio, Alain, Carine André, Bernard Teston, and Christian Cavé
2003 PERCEVAL: Une station automatisée de tests de perception et d'évaluation auditive et visuelle. *Travaux Interdisciplinaires du Laboratoire Parole et Langage* 22: 115–133.
- Kohler, Karl
1990 Macro and micro F_0 in the synthesis of intonation. In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, John Kingston and Mary E. Beckman (eds.), 115–138. Cambridge: CUP Press.
- Ladd, D. Robert, and Kim Silverman
1984 Vowel intrinsic pitch in connected speech. *Phonetica* 41: 31–40.

- Ladd, D. Robert, Dan Faulkner, Hanneke Faulkner, and Astrid Schepman
1999 Constant 'segmental' anchoring of F_0 movements under changes in speech rate. *The Journal of the Acoustical Society of America* 106(3): 1543–1554.
- Ladd, D. Robert, Ineke Mennen and Astrid Schepman
2000 Phonological conditioning of peak alignment in rising pitch accents in Dutch. *The Journal of the Acoustical Society of America* 107: 2685–2696.
- Lahiri, Aditi, and Henning Reetz
2002 In *Laboratory Phonology 7*, Carlos Gussenhoven, and Natasha Warner (eds.). Berlin/New York: Mouton de Gruyter.
- Levelt, Willem J. M.
1989 The speaker as information processor. In *Speaking: from intention to articulation*, Willem J. M. Levelt (ed.), 1–28. Cambridge: The MIT Press.
1993 Accessing words in speech production: Stages, processes and representations. In *Knowledge and language I*, Eric Reuland and Werner Abraham (eds.), 241–251. Dordrecht: Kluwer.
- Mattei, Marco, and Maria Gabriella Di Benedetto
2000 Acoustic analysis of singleton and geminate nasals in Italian. *The European Journal of Language and Speech*.
<http://www.essex.ac.uk/web-sls/>
- Moulines, Eric, and Francis Charpentier
1990 Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9: 453–467.
- Norris, Dennis, James M. McQueen, and Anne Cutler
2000 Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23: 299–325.
- Petrone, Caterina, and Mariapaola D'Imperio
(forthcoming) Effects of the interaction between tonal alignment and segmental level of speech in intonation and lexical identification.
- Peterson, Gordon, and Ilse Lehiste
1960 Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32: 693–703.
- Pierrehumbert, Janet B., and David Talkin
1991 Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, G. Jerard Docherty, and D. Robert Ladd (eds.), 90–117. Cambridge: CUP Press.
- Prieto, Pilar, Jan P. H. van Santen, and Julia Hirschberg
1995 Tonal alignment patterns in Spanish. *Journal of Phonetics* 23: 429–451.
- Rietveld, Toni, and Carlos Gussenhoven
1995 Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics* 23: 375–385.
- Scherer, Klaus R., D. Robert Ladd, and Kim Silverman
1984 Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America* 76: 1346–1356.
- Silverman, Kim E. A.
1987 The structure and processing of fundamental frequency contours. Ph. D. diss., Department of Linguistics, University of Cambridge.

- 1990 The separation of prosodies: comments on Kohler's paper. In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, John Kingston, and Mary E. Beckman (eds.), 139–151. Cambridge: CUP Press.
- Silverman, Kim E. A., and Pierrehumbert, Janet B.
 - 1990 The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, John Kingston, and Mary E. Beckman (eds.), 72–106. Cambridge: CUP.
- Stevens, Kenneth N.
 - 2002 Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America* 111: 1872–1891.
- 't Hart, Johan
 - 1991 F₀ stylization in speech: Straight lines versus parabolas. *The Journal of the Acoustical Society of America* 90 (6): 3368–3370.
- Xu, Yi
 - 1998 Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.
- Van Santen, Jan P. H., and Julia Hirschberg
 - 1994 Segmental effects on timing and height of pitch contours. In *Proceedings of the International Conference on Spoken Language Processing*, vol.2, 719–722. Yokohama, Japan.
- Van Santen, Jan P.H., and Mariapaola D'Imperio
 - 1999 Positional effects on stressed vowel duration in Standard Italian. In John Ohala (ed.) *Proceedings of the 14th International Conference of Phonetic Sciences '99* (vol.1, 241–244), San Francisco, USA.
- Wood, S.
 - 2005 Praat for beginners (User's Manual)
<http://www.ling.lu.se/persons/Sidney/praat>